BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 6: e91297
doi: 10.3897/biss.6.91297

OPEN ACCESS

Conference Abstract

# Rethinking Collection Management Data Models

Ben Collier‡, Matt Woodburn‡

‡ Natural History Museum, London, United Kingdom

Corresponding author: Ben Collier (ben.collier@nhm.ac.uk)

Received: 05 Aug 2022 | Published: 23 Aug 2022

Citation: Collier B, Woodburn M (2022) Rethinking Collection Management Data Models. Biodiversity Information Science and Standards 6: e91297. https://doi.org/10.3897/biss.6.91297

## Abstract

The data modelling of physical natural history objects has never been trivial, and the need for greater interoperability and adherence to multiple standards and internal requirements has made the task more challenging than ever. The Natural History Museum's internal RECODE (Rethinking Collections Data Ecosystems; see Dupont et al. 2022) programme has taken the approach of creating a data model to fit these internal and external requirements, rather than try and force an existing data model to work with our next generation collections management system (CMS) requirements. In this regard, community standards become vitally important, and existing and emerging standards and models like Spectrum, Darwin Core, Access to Biological Collection Data (ABCD) (Extended for Geosciences (EFG)), Latimer Core and The Conceptual Reference Model from the International Committee for Documentation (CIDOC CRM) have and will be used heavily to inform this work. The poster will provide a starting point for: publicly sharing and discussing the work that the RECODE programme has done; eliciting ideas that members of the community may have regarding its continuing improvement.

We have concentrated on creating a backbone for the data model, from collecting, through the object curation to the scientific identification. This has yielded two significant outcomes:

1. **The Collection Object:** Traditional CMS data models treat each specimen as a single record in the database. The RECODE model recognises that there are a number of different concepts that need their own entities:
    1. ***Collected material:*** the specimens collected in the field are not always fully identified or separated into discrete items.

© Collier B, Woodburn M. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

2. ***Stored object***: the aim of the RECODE model is to treat all objects as the same type of entity, with relationships between them enhancing the data. For example, a collection object is defined as a discrete object that can be moved and loaned independently. Its specific type (e.g., specimen, preparation, derivation) is given by its relationships to other collection objects.
3. ***Identifiable item***: what can be taxonomically identified does not necessarily have a 1-to-1 relationship with the stored objects. One item may contain multiple species (e.g., a parasite and host; a rock containing many minerals) or one species may be split across many objects (e.g., long branches on two or more herbarium sheets; large skeletons stored in separate locations).

2. **The Collection Level Description (CLD):** This is a construct to enable the attachment of descriptive and quantitative data to groups of collection objects, rather than individual collection object. There will always be a need for an inventory which represents the basic holdings, organisation and indexing of collections as well as a variety of use cases for grouping collection objects and attaching information at the group level.

The next challenge is to integrate the concepts more closely with each other to provide the best possible description of the collection and make it as shareable as possible. Some of the current challenges being addressed are:

- An object group may represent a heterogenous group of objects.
- There will be multiple parallel CLD schemes for different purposes.
- Different attributes and metrics will be relevant to different schemes.
- For some use cases, we need to be able to quantify relationships between an object group and its attributes as well as attaching metrics to the object group itself.
- We also need to be able to reflect relationships between object groups.

These challenges necessitate a data model that has a considerable degree of flexibility but enables rules and constraints to be introduced as appropriate for the different use cases. It is also important that, wherever possible, the model uses the same attributes as individual collection objects, to allow object groups to be implicitly linked to collection object records through common attributes as well as explicitly linked within the model. The aim of the conceptual model is to reflect these requirements.

## Keywords

RECODE, interoperability, data sharing, community engagement

## Presenting author

Ben Collier

## Presented at

TDWG 2022

## References


- Dupont S, Woodburn M, Collier B, Ellis L, Hey G, Hsu T, Sadka M, Smith D (2022) RECODE: Towards a next-gen collections management system as part of the institutional and community data ecosystem. Biodiversity Information Science and Standards https://doi.org/10.3897/biss.6.90886